Mechanism Design and Robust Control

Javier Rivas

No. 22/14

BATH ECONOMICS RESEARCH PAPERS

Department of Economics

Mechanism Design and Robust Control*

Javier Rivas
University of Bath†

May 8, 2014

Abstract

In this paper we introduce robust control methods in mechanism design problems. We assume that in the presence of an incentive compatible mechanism, players behave as if their types were in a δ-neighborhood of their true types. The designer's problem is to set up a mechanism that implements a given social choice function taking these δ-perturbations into account. In our results we characterize the social choice functions that are robust to the δ-perturbations, in the sense that the designers' loss is at most of order δ^k for a certain k. A notable finding is that in quasi-linear utilitarian environments the designer's loss is of order of δ^2.

JEL Classification: D81, D82.

Keywords: Robust Control, Mechanism Design, Bounded Rationality.

---

*I would like to thank Ludovic Renou for very useful and encouraging discussions and suggestions.

†Department of Economics, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom. j.rivas@bath.ac.uk, http://people.bath.ac.uk/fjrr20/.

1 Introduction

Consider the classic mechanism design problem of providing a public good. In this problem, the designer wants to choose how much of a public good to provide without knowing the private valuations of the beneficiaries of the public good, i.e. the players. Depending on the specifics of the problem, a mechanism usually exists such that players have incentives to truthfully reveal their true valuations to the designer, so that the designer can then choose the optimal quantity of the public good to provide the players with. In the classic mechanism design literature, such mechanism relies on the ability of players to behave rationally, i.e. players reveal their true valuations because that is how they maximize their payoffs given their belief that all other players are rational and truthful.

However, in many real life situations players may not behave fully rational. This may be for a variety of reasons, like computational constraints or memory limitations. If it can be the case that players are not rational, the designer may want to have a measure on how such irrationalities could affect the alternatives implemented by a certain mechanism. The purpose of this paper is to study how are the outcomes of the mechanism design problem affected when players are not fully rational.

We model bounded rationality by assuming that if a mechanism is in place such that players have incentives to truthfully reveal their types when all other players reveal their types (a mechanism that is incentive compatible), players may misrepresent their types. For instance, in a public good problem a player may report a valuation that is lower than his true valuation even if the mechanism in place is incentive compatible and he believes that all other players will be truthful. There are several reasons why a player may misreport his true type. It could be that he mistakenly thinks that he can "game the system" by reporting a different type, or that he simply does not fully know what his own type is.

The way players misreport their types borrows from the ideas of robust control (see, for instance, Zhou et al (1995)). We assume that for any given mechanism players behave in a rational way as if their types were somewhere in a δ-neighborhood with $\delta > 0$ of their true types. This captures the idea that the designer may be missing important information on how players behave because of their limited rationality. The designer would like to know how the alternatives implemented by each mechanism are affected by these δ-perturbations. To this end, the designer is endowed with a loss function that evaluates the differences between any two alternatives given the true types of players.

Within this context, we say that a social choice function is k-robust if the maximum loss when players misreport their types is of order δ^k with $k \geq 1$. Evidently, higher k means that the social choice function is more robust to the δ-perturbations, as the loss vanishes quickly when the size of the perturbation δ becomes small. In this paper we characterize the social choice functions that are k-robust, and obtain three main results:

First, we find that how robust a social choice function is is linked to the concept of Hölder continuity, which is a generalization of Lipschitz continuity. Not surprisingly, a robust social choice function is one where the alternatives implemented do not change dramatically when the types of players are misreported slightly. This idea of continuity is translated in terms of Hölder continuity. The usefulness of this result lies in the fact that for understanding how robust a certain social choice function is one simply has to explore its degree of Hölder continuity.

Second, we find that the only social choice functions that exhibit maximum robustness to perturbations (∞-robust social choice functions) are those that are locally constant, i.e. the alternative implemented by the social choice function is constant in the neighborhood of players' true types. There are several examples of settings where the designer may use a locally constant social choice function, from auctions to school choice problems. In all these settings the irrationality of players modeled as δ-perturbations of their reported type does not have any impact on the alternatives implemented by a mechanism. We show the applicability of our second result with an auction example.

Third, we find that in quasi-linear utilitarian environments, i.e. environments where the role of the designer is to maximize the sum of the (quasi-linear) utility of players and where the loss function is given by the differences in sums of utilities, all social choice functions are 2-robust. This means that the maximum loss caused by the δ-perturbations is of order δ^2. There are abundant examples of quasi-linear utilitarian environments in the mechanism design literature; we illustrate the usefulness of this result by means of the public good example in Bergemann and Morris (2009).

This paper combines two strands of literature. On the one hand it contributes to the analysis of robust mechanism design problems. Previous literature has looked at robust mechanism design from the perspective of the knowledge players have about the type space (Bergemann and Morris (2005)), the relationship between dominant strategies and implementation (Bergemann and Morris (2009) and Yamashita (2012)) and the designers' use of almost optimal social choice functions (Meyer-ter-Vehn and Morris (2011)). Other papers that have looked at the issue of robustness in mechanism design focus on the phenomenon of bounded rationally with adaptive players (see, for instance, Cabrales (1999) or Mathevet (2010)) or with "faulty" players (see Eliaz (2002)). Our paper differs from this strand of literature in that we look at the problem of robust mechanism design from a different angle; we study robust mechanism design using the concepts of robust control (Zhou et al (1995)).

Our paper also contributes to the strand of literature that applies the framework of robust control to economic problems. To our knowledge this literature has its roots in the work of Peters Hansen and Thomas Sargent (see Hansen and Sargent (2001), Hansen and Sargent (2007) or Williams (2008) for an overview). In contrast to these authors, we focus our efforts on mechanism design problems as opposed to macroeconomic policy problems.

The paper is organized as follows. In Section 2 we introduce the model while we present our main results in Section 3. In Section 4 we present some examples that show the applicability of our results. Finally, Section 5 concludes.

2 The Model

An environment is a tuple (N, X, Θ, u, P) where $N = \{1, \ldots, n\}$ is the set of players, X is the set of alternatives, $\Theta = \{\Theta_1, \ldots, \Theta_n\}$ where Θ_i is the set of possible types of player $i \in N$, and $u = (u_1, \ldots, u_n)$ where $u_i : X \times \Theta_i \to \mathbb{R}$ is the utility function of player $i \in N$. The function $P : \Theta \to [0, 1]$ is a probability measure on Θ and represents the common prior on the distribution of types: $P(\theta)$ is the probability that players' types are given by $\theta \in \Theta$. Each player knows his own type and player i of type $\theta_i \in \Theta_i$ holds a probabilistic belief $P(\theta_{-i}|\theta_i)$ over the types of other players $\theta_{-i} \in \Theta_{-i} = \times_{j \in N \smallsetminus \{i\}} \Theta_j$.

A social choice function $f : \Theta \to X$ associates with each profile of types $\theta \in \Theta$ an alternative $f(\theta) \in X$. Given a profile of types θ, a loss function $l_\theta : X \times X \to \mathbb{R}_+$ is a mapping from a pair of alternatives to the positive reals with $l_\theta(x, x) = 0$ for all $x \in X$ and $l_\theta(x, y) = l_\theta(y, x) \geq 0$ for all $x, y \in X$. In mathematical terms, $\{l_\theta\}_{\theta \in \Theta}$ is a collection of metrics on X where the properties of sub-additivity and the identity of indiscernibles need not be satisfied.

A mechanism is a pair (M, g) with $M = \times_{i \in N} M_i$ where M_i is player's i set of messages and $g : \times_{i \in N} M_i \to X$ is the allocation rule. An environment together with a mechanism (M, g) induce a Bayesian game $G_{(M,g)}$. A strategy profile s is given by $s = \times_{i \in N} s_i$ where s_i is the strategy of player i and it is given by $s_i : \Theta_i \to M_i$.

Let $s^*(\theta)$ be a Bayesian-Nash equilibrium of $G_{(M,g)}$ when players' types are given by $\theta \in \Theta$. The profile of strategies $s^*(\theta)$ summarizes the players' behavior that the designer considers as salient. Acknowledging that players may not be fully rational in several dimensions, the designer conjectures that each player chooses a strategy missrepreseting his type. In particular, the designer believes that players' strategies belong to the set $s^*(B_\delta(\theta))$ where $B_\delta(\theta) \subset \Theta$ is the ball of radius $\delta > 0$ around θ given some metric d. That is, according to the designer, players behave as if their types where in a δ-neighborhood of their true types.

This is how the designer acknowledges that players may not be fully rational. Players are not aware of these perturbations, they simply choose a strategy in $s^*(B_\delta(\theta))$.

Note that this paper uses mainly two sets of metrics: one set is singleton and given by the metric d on Θ and another set is given by the metrics $\{l_\theta\}_{\theta\in\Theta}$ on X. The former set of metrics is meant to specify how distances in players types are measured while the latter set of metrics specifies how differences in alternatives are evaluated by the designer. Both of these two sets of metrics are given exogenously.

An obvious alternative to the way we model players' bounded rationality is to assume that the strategies players choose lie in $B_\delta(s^*(\theta))$, i.e. players choose their strategies in the δ-neighborhood of the equilibrium strategies. Such representation of limited rationality has several conceptual problems that arise from the fact that the designer is the one that chooses the mechanism (M, g). Given that the designer is the one that chooses (M, g), he is in effect choosing the space of messages M and also its metric. Hence, the designer is indirectly choosing the set $B_\delta(s^*(\theta))$ and, thus, he can choose a metric such that $B_\delta(s^*(\theta)) = s^*(\theta)$ (which can be achieved with the discrete metric, for instance). This would mean that in effect the designer is eliminating his acknowledgment of the limited rationality of players. Given that we want to model the designers' problem when he accepts that players may be not be fully rational, he should not be allowed to ignore the inherent uncertainty that comes from such bounded rationality. In our definition of how the designer introduces the possibility of not fully rational players, the designer believes that players choose strategies as if their types where somewhere in $B_\delta(\theta)$, which is a set that follows from the set of players' types Θ and its metric d, both of which are given exogenously.

If players follow strategies s in the game $G_{(M,g)}$ and the social choice function is f, the maximum loss for a given $\theta \in \Theta$ and $\delta > 0$ is given by $\max_{\theta'\in B_\delta(\theta)} l_\theta(g(s(\theta')), f(\theta))$. Classical mechanism design sets $\delta = 0$ (i.e. $B_\delta(\theta) = \theta$) and requires that there exists a mechanism (M, g) and an equilibrium $s^*(\theta)$ of $G_{(M,g)}$ such that $g(s^*(\theta)) = f(\theta)$ for all $\theta \in \Theta$, i. e. maximum loss is zero.

We are now ready to introduce the measure of robustness used in this paper:

Definition 1. *A social choice function f is k-robustly implementable (or simply k-robust) if there exists a mechanism (M, g) such that:*

(i) The mechanism (M, g) implements f in a Bayesian-Nash equilibrium: for all $\theta \in \Theta$ there exists a Bayesian-Nash equilibrium $s^*(\theta)$ of $G_{(M,g)}$ such that $g(s^*(\theta)) = f(\theta)$.

(ii) The mechanism (M, g) bounds the maximum loss by a factor of δ^k: for all $\theta \in \Theta$ there

exist a $c > 0$ *and a* $\hat{\delta} > 0$ *such that for all* $\delta \in (0, \hat{\delta})$ *and all* $\theta' \in B_\delta(\theta)$,

$$l_\theta\left(g(s^*(\theta')), f(\theta)\right) \quad < \quad c\delta^k$$

with $k \geq 1$.

Requirement (i) of Definition 1 requires f to be implementable in the classical sense, and it implies that the social choice function is Bayesian incentive compatible: for each player i and for each type θ_i we have that

$$\int_{\theta_{-i}\in\Theta_{-i}} u_i(f(\theta_i, \theta_{-i}), \theta_i) dP(\theta_{-i}|\theta_i) \quad \geq \quad \int_{\theta_{-i}\in\Theta_{-i}} u_i(f(\theta_i', \theta_{-i}), \theta_i) dP(\theta_{-i}|\theta_i),$$

for all $\theta_i' \in \Theta_i$. In particular, truth-telling is an equilibrium of the direct revelation mechanism (Θ, f).

Requirement (ii) states that if a social choice function is k-robust then the loss due to the bounded rationality of players cannot be of a factor greater than δ^k. Note that effectively the designer is following a maxmin approach. In particular, k-robustnes implies that

$$\max_{\theta'\in B_\delta(\theta)} l_\theta\left(g(s^*(\theta')), f(\theta)\right) \quad < \quad c\delta^k,$$

and the designer would like to know what is the minimum δ^k (maximum k) such the inequality above is true. The maxmin approach is in line with previous economic models using the robust control framework (see Hansen and Sargent (2001) and the reference to Gilboa and David Schmeidler (1989) herein).

Given that the purpose of this paper is to study mechanism design in the presence of bounded rational players, we focus on social choice functions that are implementable in the classical sense. That is, we only consider social choice functions that satisfy condition (i) in Definition 1. There are several assumptions that could be made on f that guarantee that the social choice function can be implemented. Here we do not assume any conditions in particular but simply that f is implementable by some mechanism.

3 Results

Before we dwell into the characterization of k-robust social choice functions, the following result, which is a revelation principle type of result, allows us to focus only on direct mechanism where players truthfully report their types.[1]

[1]Note that players truthfully reporting types simply means that $s(\theta) = \theta$, yet in general $s(B_\delta(\theta)) \neq s(\theta)$.

Proposition 1. *A social choice function f is k-robust with mechanism (M, g) if and only if it is k-robust with mechanism (Θ, f).*

Proof. If f is k-robust with mechanism (Θ, f) then it is trivially k-robust with some mechanism (M, g), simply set $(M, g) = (\Theta, f)$.

To prove the other direction of the implication, first note that condition (i) of the definition of k-robust is always satisfied by any social choice function that is implementable.[2] Thus, we are left to prove condition (ii) of the definition of k-robustness.

If f is k-robust with mechanism (M, g) then we have that for all $\theta \in \Theta$ there exists a $c > 0$ and a $\hat{\delta} > 0$ such that for all $\delta \in (0, \hat{\delta})$ and all $\theta' \in B_\delta(\theta)$

$$l_\theta \left(g(s^*(\theta')), f(\theta) \right) \quad < \quad c\delta^k$$

Given that f is k-robust with a mechanism (M, g), using condition (i) of the definition of k-robustness it is true that for any $\theta' \in B_\delta(\theta)$ we have that $g(s^*(\theta')) = f(\theta')$ and, hence,

$$l_\theta \left(g(s^*(\theta')), f(\theta) \right) \quad = \quad l_\theta \left(f(\theta'), f(\theta) \right).$$

Therefore, combining the two expressions above:

$$l_\theta \left(f(\theta'), f(\theta) \right) \quad < \quad c\delta^k$$

for all $\theta' \in B_\delta(\theta)$.

Thus, in the direct mechanism (Θ, f) where players truthfully report their types θ and these are perturbed to $\theta' \in B_\delta(\theta)$ the loss is bounded by $c\delta^k$. $\square$

A direct consequence of the revelation principle stated above is that when looking for a characterization of k-robust social choice functions we can restrict our attention to the study of truth-telling direct mechanisms.

Corollary 1. *A social choice function f is k-robust if and only if and for all $\theta \in \Theta$ there exists a $c > 0$ and a $\hat{\delta} > 0$ such that for all $\delta \in (0, \hat{\delta})$ and all $\theta' \in B_\delta(\theta)$,*

$$l_\theta(f(\theta'), f(\theta)) < c\delta^k.$$

Therefore, Corollary 1 implies that if a certain mechanism allows f to satisfy the definition of k-robustness for a given k then the direct mechanism (Θ, f) also allows the loss to be bounded by a factor of δ^k (with the same constant c). An implication of this is that the

[2] We assume that f is implementable, see the previous section.

designer does not choose a difference mechanism than the one he would choose if he ignored the limited rationality of players. This is an implication of the condition (i) in the definition of k-robustness. Relaxing condition (i) in the definition of k-robustness could allow the designer to choose mechanism that, although not being incentive compatible, may create a smaller loss for some values of δ than an incentive compatible mechanism. We believe that incentive compatibility must be a requirement on the social choice function because players may still behave according to their true types (this is a possibility with the δ-perturbations), and as such the mechanism must be able to implement the same alternatives it was first set out to implement in the absence of bounded rationality.

Once we have established that the presence players' limited rationality does not make the designer to change the mechanism he chooses, a question that arises is that of the characterization of the loss induced by the δ-perturbations, i.e. the characterization of k-robust social choice functions for all k. Our analysis continues by exploring this issue.

One might guess that k-robust social choice functions should exhibit some type of continuity, so that a perturbation of order δ is not translated into losses of an order much grater than δ. This continuity is present in terms of Hölder continuity, which is a generalization of Lipschitz continuity:

Definition 2. *A social choice function f is locally Hölder continuous of degree k if for any $\theta \in \Theta$ there exists a $c > 0$ and a $\hat{\delta} > 0$ such that for all $\delta \in (0, \hat{\delta})$ and all $\theta' \in B_\delta(\theta)$*

$$l_\theta(f(\theta'), f(\theta)) \leq c d(\theta', \theta)^k.$$

We have the following result:

Proposition 2. *A social choice function f is k-robust if and only if it is locally Hölder continuous of degree k.*

Proof. Take any $\theta \in \Theta$. If f is k-robust then there exists a $c > 0$ and a $\hat{\delta} > 0$ such that for all $\delta \in (0, \hat{\delta})$ and all $\theta' \in B_\delta(\theta)$

$$l_\theta\left(f(\theta')), f(\theta)\right) \quad < \quad c\delta^k.$$

Assume that there exists a $\bar{\theta} \in B_\delta(\theta)$ such that

$$l_\theta(f(\bar{\theta}), f(\theta)) \quad > \quad c d(\bar{\theta}, \theta)^k.$$

Then, there exists a $\bar{\varepsilon} > 0$ such that for all $\varepsilon \in (0, \bar{\varepsilon})$ if we define $\bar{\delta} = d(\bar{\theta}, \theta) + \varepsilon$ we have that $\bar{\delta} < \hat{\delta}$ and, since $\bar{\delta} > d(\bar{\theta}, \theta)$, also that $\bar{\theta} \in B_{\bar{\delta}}(\theta)$. Thus, because f is k-robust we obtain that

$$
\begin{aligned}
l_\theta\left(f(\bar{\theta})), f(\theta)\right) &< c\bar{\delta}^k \\
&< c(d(\bar{\theta},\theta)+\varepsilon)^k.
\end{aligned}
$$

Since the inequality above is true for all $\varepsilon \in (0,\bar{\varepsilon})$:

$$
l_\theta\left(f(\bar{\theta})), f(\theta)\right) \leq cd(\bar{\theta},\theta)^k,
$$

which represents a contradiction. Therefore, it is true that for all $\theta \in \Theta$ there exists a $c > 0$ and a $\hat{\delta} > 0$ such that for all $\delta \in (0,\hat{\delta})$ and all $\theta' \in B_\delta(\theta)$

$$
l_\theta\left(f(\theta')), f(\theta)\right) \leq cd(\theta',\theta)^k.
$$

This is the definition of Hölder continuity of degree k.

Assume now that f is locally Hölder continuous of degree k. We have then that for any $\theta \in \Theta$ there exists a $c > 0$ and a $\hat{\delta} > 0$ such that for any $\delta \in (0,\hat{\delta})$ and any $\theta' \in B_\delta(\theta)$

$$
l_\theta(f(\theta'), f(\theta)) \leq cd(\theta',\theta)^k.
$$

Since $d(\theta',\theta) < \delta$ we have that

$$
l_\theta(f(\theta'), f(\theta)) < c\delta^k
$$

as required. □

Note that the definition of Hölder continuity is made with respect to the metrics $\{l_\theta\}_{\theta\in\Theta}$, i.e. taking f to be a function between the two metric spaces $(\Theta, d) \to (X, l_\theta)$ where $\theta \in \Theta$ is the true type of players. Thus, it could be that if, for instance, $X = \mathbb{R}$, then f is locally Hölder continuous as a mapping $(\Theta, d) \to (X, l_\theta)$ but not as a mapping $(\Theta, d) \to (X, E)$ where E is the euclidean distance. The metrics $\{l_\theta\}_{\theta\in\Theta}$ measure how far apart in terms of loses for the designer two different alternatives are while the Euclidean distance measures how far apart in space two different alternatives are. Thus, if alternatives are compared using different metrics then the fact that f is Hölder continuous with respect to one metric does not imply that it is also Hölder continuous with respect to another metric. This situation arises in the example in Section 4.2.

Proposition 2 states a full characterization of social choice functions. For knowing whether a social choice function is k-robust or not it is sufficient to study its degree of Hölder continuity. As we already mentioned, the fact that the notion of k-robustness is linked with a certain type of continuity is not surprising, as by definition the concept of robustness incorporates the idea that small perturbations in the given parameters should not lead to big changes in the alternatives selected by the social choice function.

Next, we focus on social choice functions for which the limited rationality of players creates no loss (i.e. the set of ∞-robust social choice functions). Before we do that, however, a new definition is in order:

Definition 3. *A social choice function f is locally constant if for all θ there exists a $\hat{\delta} > 0$ such that for all $\delta \in (0, \hat{\delta})$ and all $\theta' \in B_\delta(\theta)$ we have that $l_\theta(f(\theta'), f(\theta)) = 0$.*

Remark 1. *A social choice function f is locally constant if and only if it is locally Hölder continuous of degree ∞.*

A consequence of the remark above and Proposition 2 is the following result:

Corollary 2. *A social choice function f is ∞-robust if and only if it is locally constant.*

Locally constant social choice functions are frequent in the social choice and mechanism design literature. As we shall see in Section 4.1, examples of locally constant social choice functions appear in settings where there is an indivisible object to share amongst some claimants (i.e. auctions, the Solomon's Dilemma, etc.). These settings are characterized by the fact that small perturbation in player's types do not lead to changes in who the social choice function allocates the object to. Corollary 2 implies that social choice functions in these environments are ∞-robust.

3.1 Quasi-linear Utilitarian Environments

We continue the study of k-robust social choice functions by imposing some structure in the environments we deal with. In particular, in this section we focus on quasi-linear utilitarian environments: environments where the goal of the designer is to choose an allocation that maximizes the the aggregate sum of the utility of all players, and where the utility of players is linear in wealth. Specifically, we assume the following:

Assumption 1. *Alternatives are of the form $x = (x_0, x_1, \ldots, x_n) \in X = \mathbb{R}^{n+1}$ where x_0 represents a certain choice of the designer and x_i with $i \in N$ are the transfers of each player. Moreover,*

$$u_i(x, \theta_i) \quad = \quad v_i(x_0, \theta_i) - \sum_{j=1}^{n} a_j x_j$$

with $v_i : \mathbb{R} \times \Theta_i \to \mathbb{R}$ and $a_j \in \mathbb{R}$ for all $i, j \in N$.

The next assumption specifies how the social choice function selects among different alternatives and how the designer evaluates losses.

Assumption 2. *Define* $e = (1, 0 \ldots, 0)$*. For all* $\theta \in \Theta$ *and all* $x, y \in X$*:*

- $ef(\theta) = \arg\max_{x_0 \in \mathbb{R}} \sum_{i=1}^{n} v_i(x_0, \theta_i)$,
- $l_\theta(x, y) = |\sum_{i=1}^{n} v_i(x_0, \theta_i) - \sum_{i=1}^{n} v_i(y_0, \theta_i)|$.

That is, the designer would like to choose x_0 in order to maximize the sum of the utilities of all players ignoring the transfers.[3]. On top of that, the loss function measures the welfare difference between two alternatives, quantified as the difference in the sum of the utility of x_0 for each player.

We assume that in quasi-linear utilitarian environments the designer selects the optimal x_0 using a first order condition, i.e. ef maximizes the sum of utilities by differentiating and setting the derivative equal to zero:[4]

Assumption 3. *For all* $\theta \in \Theta$*,*

$$\frac{\partial \sum_{i=1}^{n} v_i(ef(\theta), \theta_i)}{\partial ef(\theta)} = 0.$$

We also impose some structure on the utility functions and the social choice functions. Firstly, we assume that for all players' types $\theta \in \Theta$ the term $\sum_{i=1}^{n} v_i$ coincides with its analytic form (Taylor expansion) around $ef(\theta)$. In particular:

Assumption 4. *For all* $\theta \in \Theta$ *and all* $x_0 \in \mathbb{R}$ *the term* $\sum_{i=1}^{n} v_i$ *is such that:*

$$\sum_{i=1}^{n} v_i(x_0, \theta_i) = \sum_{j=0}^{\infty} \frac{\partial^j \sum_{i=1}^{n} v_i(ef(\theta), \theta_i)}{\partial^j ef(\theta)} \frac{(x_0 - ef(\theta))^j}{j!},$$

where there exists an upper bound $r > 0$ *such that*

$$r > \sup_{\theta \in \Theta, j \in \mathbb{N}^+} \left| \frac{\partial^j \sum_{i=1}^{n} v_i(ef(\theta), \theta_i)}{\partial^j ef(\theta)} \frac{1}{j!} \right|.$$

Assumption 4 is satisfied if, for instance, the function $\sum_{i=1}^{n} v_i$ is a polynomial with finite coefficients.

Finally, we assume that the choice of the designer, ef, is Lipschitz continuous:

[3]Transfers are only used to allow the mechanism employed by the designer to be incentive compatible. For instance, in the public good example we present in Section 4.2 transfers are used to finance the public good.

[4]Implicit in this assumption is the fact that such partial derivative of $\sum_i v_i$ exists.

Assumption 5. *There exists a $k > 0$ such that for any $\theta, \theta' \in \Theta$ we have that*

$$\left| ef(\theta') - ef(\theta) \right| \quad \leq \quad kd(\theta', \theta).$$

As we shall see, the structure imposed by Assumptions 1-5 is in line with settings commonly found in the literature. An instance of these settings is presented is Section 4.2 where we deal with the public good example found in Bergemann and Morris (2009).

Note that the fact that ef is Lipschitz continuous (Assumption 5) does not imply that f is locally Hölder continuous of degree 1. This is because of two reasons. First, Assumption 5 only deals with the first component of f. Second, the definition of Lipschitz continuity is made with respect to the Euclidean distance while the definition of Hölder continuity is made with respect to the metrics $\{l_\theta\}_{\theta\in\Theta}$. In loose terms, in our setting the Euclidean metric measures how far apart in space two different alternatives are while the metrics $\{l_\theta\}_{\theta\in\Theta}$ measure how far apart in terms of loses for the designer two different alternatives are. For utilitarian environments, the only assumption we make about how the designer evaluates loses is Assumption 2.

We have the following characterization:

Proposition 3. *All social choice functions in quasi-linear utilitarian environments (Assumptions 1-5) are 2-robust.*

Proof. We have that in a quasi-linear utilitarian environment for all $\theta', \theta \in \Theta$:

$$\begin{aligned}
l_\theta(f(\theta'), f(\theta)) &= \left| \sum_{i=1}^{n} \left(v_i(ef(\theta'), \theta_i) - v_i(ef(\theta), \theta_i) \right) \right| \\
&= \left| \sum_{j=0}^{\infty} \frac{\partial^j \sum_{i=1}^n v_i(ef(\theta)), \theta_i)}{\partial^j ef(\theta)} \frac{(ef(\theta') - ef(\theta))^j}{j!} - \sum_{i=1}^{n} v_i(ef(\theta), \theta_i) \right| \\
&= \left| \sum_{j=1}^{\infty} \frac{\partial^j \sum_{i=1}^n v_i(ef(\theta), \theta_i)}{\partial^j ef(\theta)} \frac{(ef(\theta') - ef(\theta))^j}{j!} \right|.
\end{aligned}$$

Combining the fact that $\frac{\partial \sum_{i=1}^n v_i(ef(\theta),\theta_i)}{\partial ef(\theta)} = 0$ (Assumption 3) with the upper bound r (Assumption 4) with gives:

$$\begin{aligned}
l_\theta(f(\theta'), f(\theta)) &= \left| \sum_{j=2}^{\infty} \frac{\partial^j \sum_{i=1}^n v_i(ef(\theta), \theta_i)}{\partial^j ef(\theta)} \frac{(ef(\theta') - ef(\theta))^j}{j!} \right| \\
&\leq r \sum_{j=2}^{\infty} \left| ef(\theta') - ef(\theta) \right|^j .
\end{aligned}$$

By Assumption 5 there exists a $k > 0$ such that for all $\theta \in \Theta$ and all $\delta > 0$ we have that for all $\theta' \in B_\delta(\theta)$

$$\begin{aligned} \left| ef(\theta') - ef(\theta) \right| &\leq kd(\theta', \theta) \\ &< k\delta. \end{aligned}$$

Choose some $\hat{\delta} \in (0, \frac{1}{k})$ such that $(k\delta)^2 > \sum_{j=3}^{\infty}(k\delta)^j$ for all $\delta \in (0, \hat{\delta})$ (for example, $\hat{\delta} < \frac{1}{2k}$). We have that for all $\theta \in \Theta$ and all $\theta' \in B_\delta(\theta)$ with $\delta \in (0, \hat{\delta})$:

$$\begin{aligned} l_\theta(f(\theta'), f(\theta)) &< r\sum_{j=2}^{\infty}(k\delta)^j \\ &< 2rk^2\delta^2. \end{aligned}$$

Therefore, f is 2-robust. $\square$

Proposition 3 shows that all social choice function in quasi-linear utilitarian environments are 2-robust. The intuition for this result is that if the alternative to be implemented by the social choice function is calculated with a first order condition, then infinitesimal changes in the alternative chosen do not change the value of the objective function (derivative equals zero). Hence, if the types of players are perturbed slightly and the alternative implemented does not change much as a result (ef is Lipschitz continuous), the first order effect of this perturbation (the term of order δ) is zero, and only the second order effect (the term of order δ^2) matters.

4 Illustration of Results

In this section we illustrate the applicability of our main results with two examples. The first example deals with an auction setting where the designer wants to allocate an object to the bidder that values it the most. This example shows an application of Corollary 2. The second example illustrates an application of the result in Proposition 3 in a quasi-linear utilitarian environment where the role of the designer is to choose how much of a public good to provide.

4.1 Single Unit Auction

Consider an auction where two bidders $N = \{1, 2\}$ compete for an indivisible good. The set of alternatives is $X = N \times \mathbb{R}$ where the alternative $x = (i, p)$ represents the situation where player i takes the item paying a price of p. The player who wins the auction on each allocation x is referred to as x_W. Each bidder $i \in \{1,2\}$ values the object at $\theta_i \in \mathbb{R}$ (where $\mathbb{R}$ is endowed the euclidean distance) and has a utility is given by the valuation of the item minus the price he pays in case he wins the auction. That is, $u_i(x,\theta_i) = (\theta_i - p)\mathbb{1}_{x_W = i}$.

The social choice function is given by $f(\theta) = (1,p)$ if $\theta_1 > \theta_2$ and $f(\theta) = (2,p)$ if $\theta_2 > \theta_1$ for any $p \leq \max\{\theta_1, \theta_2\}$. We are ignoring the case where $\theta_1 = \theta_2$ as in this case it is irrelevant which bidder wins the auction. This social choice function is implementable by, for instance, the second price auction.

The loss function is given by $l_\theta(x,y) = h(|\theta_1 - \theta_2|)\mathbb{1}_{x_W \neq y_N}$ where $h : \mathbb{R}_+ \to \mathbb{R}_+$ is weakly increasing with $h(0) = 0$. That is, if both $x, y \in X$ prescribe the same winner of the auction then the loss is zero, otherwise the loss is increasing in the differences in types.

Note that in the environment just defined the social choice function is locally constant. Indeed, for all $\theta \in \Theta$ if we set $\hat{\delta} = \frac{|\theta_1 - \theta_2|}{2}$ then for all $\delta \in (0, \hat{\delta})$ and for all $\theta' \in B_\delta(\theta)$ we have that $\theta_1 - \theta_2 > 0$ implies $\theta'_1 - \theta'_2 > 0$: if $\theta_1 - \theta_2 > 0$ then $\theta'_1 - \theta'_2 > \theta_1 - \delta - (\theta_2 + \delta) = |\theta_1 - \theta_2| - 2\delta > 0$. Similarly, $\theta_1 - \theta_2 < 0$ implies $\theta'_1 - \theta'_2 < 0$: if $\theta_1 - \theta_2 < 0$ then $\theta'_1 - \theta'_2 < \theta_1 + \delta - (\theta_2 - \delta) = -|\theta_1 - \theta_2| + 2\delta < 0$. Hence, the allocation when players missreport their types to $\theta' \in B_\delta(\theta)$ is the same allocation as when they report their true types. Thus, for all $\theta \in \Theta$ and all $\delta \in \left(0, \frac{|\theta_1 - \theta_2|}{2}\right)$ we have that $l_\theta(f(\theta), f(\theta')) = 0$ for all $\theta' \in B_\delta(\theta)$.

Remark 2. *The social choice function f in the Single Unit Auction is locally constant and, hence, by Corollary 2 it is ∞-robust.*

In this example the fact that the optimal allocation is locally constant implies that slight changes in the types of players will not change the identity of the bidder who values the item the most amongst the two bidders. Hence, if the perturbations to players' types are insignificant enough, i.e. δ is small enough, then the second price auction still allocates the item to the bidder that values the item the most.

4.2 Provision of a Public Good

Consider the public good example in Bergemann and Morris (2009) with no interdependent utility functions.[5] The set of players is $N = \{1,2,3\}$ and the set of possible allocations in this setting is given by $X = \mathbb{R}_+ \times \mathbb{R}^3$ where if $x = (x_0, x_1, x_2, x_3)$ then x_0 units of the public good are provided and the contribution of agent $i \in \{1,2,3\}$ is given by x_i. The cost of providing an amount x_0 of the public good is given by $\frac{1}{2}x_0^2$. The objective of the designer is to choose $x = (x_0, x_1, x_2, x_3)$ such that the sum of the utility of all players is maximized. The player's

[5]This means that the parameter γ in Bergemann and Morris (2009) is set to 0.

types are given by $\Theta = \times_{i=1}^{3}\Theta_i$ where for $i = \{1,2\}$ we have that $\Theta_i = \mathbb{R}$ is endowed with the Euclidean distance and $\Theta_3 = \{0\}$. Note that for any $\theta \in \Theta$, any $\delta > 0$ and any $\theta' \in B_\delta(\theta)$ it is true that $\theta'_3 = \theta_3$.

Player's $i \in \{1,2\}$ utility function is given by

$$u_i\,(x,\theta_i) \;=\; \theta_i x_0 - x_i$$

and player's 3 utility function is given by

$$u_3\,(x,\theta_3) \;=\; -\frac{1}{2}x_0^2 + x_1 + x_2.$$

Note that player 3 simply represents the designer's wealth.[6] Using the notation from Section 3.1 we have that $v_i\,(x,\theta_i) = \theta_i x_0$ for $i \in \{1,2\}$ and $v_3\,(x,\theta_3) = -\frac{1}{2}x_0^2$. Note that the environment just defined satisfies Assumption 1.

The designer chooses

$$x_0 \;=\; \arg\max_{x_0} \theta_1 x_0 + \theta_2 x_0 - \frac{1}{2}x_0^2$$

and, hence, we have that $x_0 = \theta_1 + \theta_2$. As argued in Bergemann and Morris (2009) this amount of public good is implementable by the Vickrey-Clarke-Groves transfers given by $x_i = -\frac{1}{2}\theta_i^2$ for $i \in \{1,2\}$. Thus, the social choice function is given by $f(\theta) = (\theta_1+\theta_2, \frac{1}{2}\theta_1^2, \frac{1}{2}\theta_2^2, \frac{1}{2}\theta_1^2 + \frac{1}{2}\theta_2^2)$.

We assume that the designer evaluates loses according to the loss function $l_\theta(x,y) = |\sum_{i=1}^{n} v_i(x_0,\theta_i) - \sum_{i=1}^{n} v_i(y_0,\theta_i))\,| = \left|(\theta_1+\theta_2)x_0 - \frac{1}{2}x_0^2 - \left((\theta_1+\theta_2)y_0 - \frac{1}{2}y_0^2\right)\right|$. Therefore, since $x_0 = ef(\theta)$ maximizes $\sum_{i=1}^{n} v_i(x_0,\theta_i) = (\theta_1+\theta_2)x_0 - \frac{1}{2}x_0^2$, the environment just defined satisfies Assumption 2.

Under mechanism (Θ, f), if players report types θ then we have that $x_0 = ef(\theta)$ and, hence, we can write $\sum_{i=1}^{n} v_i(ef(\theta),\theta_i) = (\theta_1+\theta_2)ef(\theta) - \frac{1}{2}(ef(\theta))^2$. A consequence of this is that $\frac{\partial \sum_{i=1}^{n} v_i(ef(\theta),\theta_i)}{\partial ef(\theta)} = 0$. Thus, Assumption 3 is also satisfied.

Moreover, it is true that $\frac{\partial^2 \sum_{i=1}^{n} v_i(ef(\theta),\theta_i)}{\partial^2 ef(\theta)} = -1$ and $\frac{\partial^2 \sum_{i=1}^{n} v_i(ef(\theta),\theta_i)}{\partial^j ef(\theta)} = 0$ for all $j \geq 3$. Hence, for all $\theta' \in \Theta$:

$$\sum_{j=0}^{\infty} \frac{\partial^j \sum_{i=1}^{n} v_i(ef(\theta),\theta_i)}{\partial^j ef(\theta)} \frac{|ef(\theta') - ef(\theta)|^j}{j!} \;=\; (\theta_1+\theta_2)ef(\theta) - \frac{1}{2}(ef(\theta))^2 - \frac{|ef(\theta') - ef(\theta)|^2}{2}.$$

Since

$$\left|ef(\theta') - ef(\theta)\right| \;=\; \left|(\theta'_1 + \theta'_2) - (\theta_1+\theta_2)\right|,$$

[6] In Bergemann and Morris (2009) player 3 is not needed. We introduce player 3 so that we can apply our main results in this example but player 3 plays no role in the strategic incentives of the other players nor in those of the designer: player 3 does not affect the social choice function nor the transfers, and his reported type has no effect on the alternative that is implemented.

we have that

$$\begin{aligned}
\sum_{j=0}^{\infty} \frac{\partial^j \sum_{i=1}^n v_i(ef(\theta),\theta_i)}{\partial^j ef(\theta)} \frac{|ef(\theta')-ef(\theta)|^j}{j!} &= \frac{1}{2}(\theta_1+\theta_2)^2 - \frac{|(\theta_1'+\theta_2')-(\theta_1+\theta_2)|^2}{2} \\
&= (\theta_1+\theta_2)(\theta_1'+\theta_2') - \frac{1}{2}(\theta_1'+\theta_2')^2 \\
&= \sum_{i=1}^n v_i(ef(\theta'),\theta_i).
\end{aligned}$$

Therefore, Assumption 4 is satisfied. Finally, since for all $\theta' \in B_\delta(\theta)$ we have that $|ef(\theta')-ef(\theta)| \leq |\theta_1'-\theta_1|+|\theta_2'-\theta_2| \leq 2\delta$, the social choice function f is Lipschitz continuous. Hence, Assumption 5 is also satisfied. This implies that Assumptions 1-5 are satisfied and by Proposition 3 we have that f is 2-robust.

Remark 3. *The environment just defined is quasi-linear utilitarian and, hence, by Proposition 3 the social choice function f is 2-robust.*

Next we confirm this fact by explicitly calculating the loss when the players report a type different than their true types. For all $\theta \in \Theta$ if we fix any $\hat{\delta} > 0$ then for all $\delta \in (0,\hat{\delta})$ and all $\theta' \in B_\delta(\theta)$ we have that

$$\begin{aligned}
l_\theta(\theta',\theta) &= \left| \sum_{i=1}^3 v\left(ef(\theta'),\theta_i\right) - \sum_{i=1}^3 v_i\left(ef(\theta),\theta_i\right) \right| \\
&= \left| (\theta_1+\theta_2)(\theta_1'+\theta_2') - \frac{1}{2}(\theta_1'+\theta_2')^2 - \frac{1}{2}(\theta_1+\theta_2)^2 \right| \\
&= \left| (\theta_1+\theta_2)(\theta_1'+\theta_2') - \frac{1}{2}\left[(\theta_1'+\theta_2')^2 - (\theta_1+\theta_2)^2\right] - (\theta_1+\theta_2)^2 \right| \\
&= \left| (\theta_1+\theta_2)\left[(\theta_1'+\theta_2') - (\theta_1+\theta_2)\right] - \frac{1}{2}\left[(\theta_1'+\theta_2') - (\theta_1+\theta_2)\right]\left[(\theta_1'+\theta_2') + (\theta_1+\theta_2)\right] \right| \\
&= \frac{1}{2}\left[(\theta_1'+\theta_2') - (\theta_1+\theta_2)\right]^2 \\
&= \frac{1}{2}\left[(\theta_1'-\theta_1) + (\theta_2'-\theta_2)\right]^2 \\
&< \frac{1}{2}(2\delta)^2 \\
&< 2\delta^2
\end{aligned}$$

Thus, as already anticipated, f is 2-robust.

5 Conclusions

This paper investigates bounded rationality in mechanism design problems. Bounded rationality is modeled borrowing from the ideas in the literature on robust control; we assume that for a given mechanism players behave as if their types were in a δ-neighborhood of their true types. The designer acknowledges this fact and would like to know how the alternatives chosen by each mechanism are affected by these δ-perturbations. To this end, he is endowed with a loss function that evaluates the differences between any two alternatives given the true types of players. We say that a social choice function is k-robust if the maximum loss when players misreport their types is of order δ^k.

In our results we obtain three main conclusions. First, we find that a social choice function is k-robust if and only if it is Hölder continuous of degree k. Second, we find that the only social choice functions that exhibit maximum robustness to perturbations are those that are locally constant. Third, we find that all social choice functions in quasi-linear utilitarian environments are 2-robust.

Our results offer new insights on how small perturbations may affect the alternatives chosen by a given mechanism. We include two illustrations in the paper in order to highlight the applicability of our results. To our knowledge, our paper is the first one to study robust mechanism design using the tools of robust control.

References

- Bergemann, D. and S. Morris (2005): "Robust Mechanism Design", *Econometrica* 73 (6), 1771-1813.

- Bergemann, D. and S. Morris (2009): "Robust Implementation in Direct Mechanisms", *The Review of Economic Studies* 76, 1175-1204.

- Cabrales, A. (1999): "Adaptive Dynamics and the Implementation Problem with Complete Information", *Journal of Economic Theory* 86, 159-184.

- Eliaz, K. (2002): "Fault Tolerant Implementation", *The Review of Economic Studies* 69 (3), 589-610.

- Gilboa, I. and D. Schmeidler (1985): "Maxmin Expected Utility with Non-unique Prior", *Journal of Mathematical Economics* 18, 141-153.

- Hansen, L. P. and T. J. Sargent (2001): "Robust Control and Model Uncertainty", *The American Economic Review Papers and Proceedings* 91 (2) 60-66.

- Hansen, L. P. and T. J. Sargent, Robustness, Princeton University Press, 2007.

- Mathevet, L. (2010): "Supermodular mechanism design", *Theoretical Economics* 5, 403-443.

- Meyer-ter-Vehn, M. and S. Morris (2011): “The robustness of robust implementation”, *Journal of Economic Theory* 146, 2093-2104.

- Williams, N., *Robust Control* in The New Palgrave Dictionary of Economics, 2008.

- Yamashita, T. (2012): ‘A Necessary Condition for Implementation in Undominated Strategies, with Applications to Robustly Optimal Trading Mechanisms” working paper.

- Zhou, K. J. C. Doyle and K. Glover, Robust and Optimal Control, Prentice Hall, 1995.